Breakout 2B
Archives/Backups as extensions to file systems

Session Coordinators: Grider

Session Scribes: Miller

Session Presenter: Grider

Session Writeup:

Each Breakout session will provide

1) Current high level topics of File Systems and I/O Research in this area

2) Areas that need to have more research focus

3) Areas that have or will have too much research focus

4) Some rough consensus ranking of areas that need more focus,
less focus,
and overall recommendations including
Short term research needs
Long term research needs

There will be a presentation of this material for each session done by the session leader and a write up for inclusion in the workshop documentation.

Current high level topics of File Systems and I/O Research in this area

- *Object Archive*
- *File system on the tape, read ahead, giving user direct control of tape via file system interface*
- *Deep archive on disk instead of tape, dealing with overhead/failure*
- *Content addressable archiving/restoration*
- *Tape emulation, VTL, Maid*
- *Schedule and managing data movement over life time*
- Continuous versioning

Areas that need to have more research focus (designate short and long term)

1. Object archives (MED), partial retrieve, etc.
2. POSIX as a users interface to archive (?)
3. DMAPI, is it dead, should it be, maybe object archives makes it irrelevant (?)
4. Is tape speed keeping up with density of tape, what problems does this cause for things like technology insertion
5. Archive metadata management and long term migration and continuity
6. Archive metadata standardization which includes extensible, and open research into how to standardize describing media so they can be moved forward to a new archive
7. Is the access and security model different for archives versus file systems.
8. Scale of the metadata (size and ops)
9. Parallel archiving
10. *File system on the tape, read ahead, giving user direct control of tape via file system interface*
11. *Deep archive on disk instead of tape, dealing with overhead/failure*
12. *Content addressable archiving/restoration*
13. *Tape emulation, VTL, Maid*
14. *Schedule and managing data movement over life time*
15. Continuous versioning
16. Key management over the long period for archive data

Areas that have or will have enough or too much research focus (designate short and long term)

- None

Some rough consensus ranking of areas that need more focus, less focus and overall recommendations including Short/Long research needs

1. Archive metadata management, extensibility, continuity, standardization (to move from system to system over time) and open research into how to standardize fully describing media formats so they can be moved forward to a new archive system over time (transparency)
    1. Total 74 Government 22
2. Deep archive on disk instead of tape, dealing with implications like failure etc., content addressable archiving/restoration, and continuous versioning
    1. Total 48 Government 16 med term
3. Object archives, parallel archiving, and partial retrieval (may be combined with 4)
    1. Total 32 Government 19 med term
4. POSIX as a users interface to archive, replacement for DMAPI function (may be combined with 3)
    1. Total 27 Government 18 med-long term
5. Scheduling and managing data movement over life time (non combined)
    1. Total 20 Government 10 med-long term
6. Key management over the long period for archive data
    1. Total 19 Government 2 <- disagreement
    2. .edu and .com like this one med-long term
    3. It exists elsewhere (it is in the security breakout as number 2)